هوش مصنوعی
تمرین نوشتاری شماره ۱

مدرس: **دکتر فدایی**

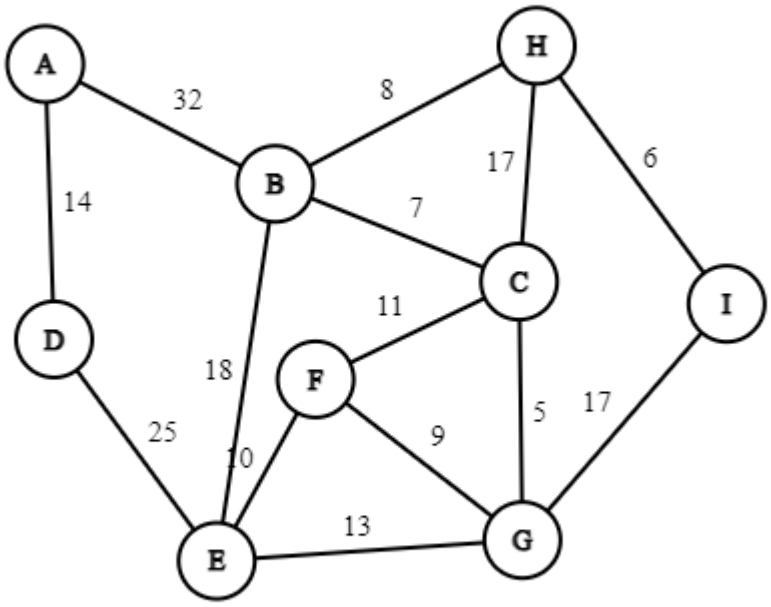
طراحان: امید پناکاری، پویا صادقی، علی اخگری، مجید فریدفر

مهلت تحویل: یکشنبه ۲۱ آبان ۱۴۰۲، ساعت ۲۳:۵۹

Search

سوال اول

در گراف زیر می‌خواهیم با کوتاهترین مسیر از راس A به I را بیابیم.



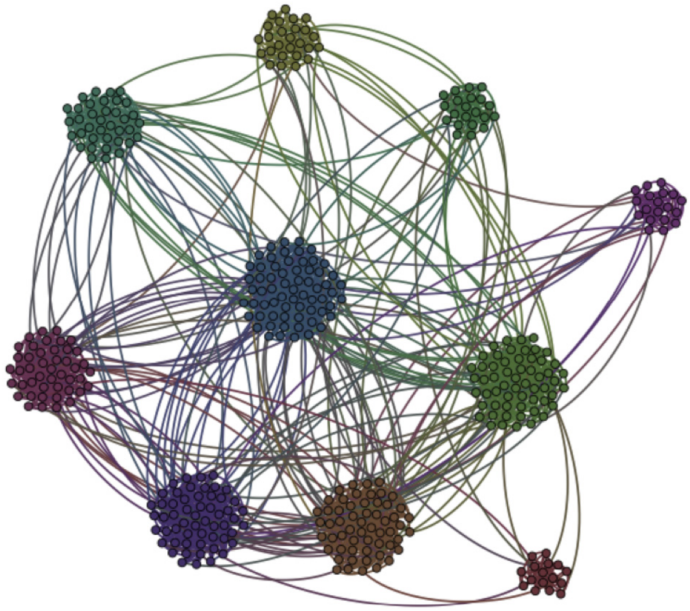
با استفاده از الگوریتم Uniform Cost Search کوتاهترین مسیر از A به I را بیابید. در صورت وجود چندین گزینه در یک مرحله، راس کوچکتر از نظر الفبایی را انتخاب کنید. در طی تمام مراحل مجموعه frontier، مجموعه explored و آخرین راس اضافه شده به explored در آن مرحله را به همراه مسیر طی شده و هزینه مصرف شده تا آن راس را بنویسید.

سوال دوم

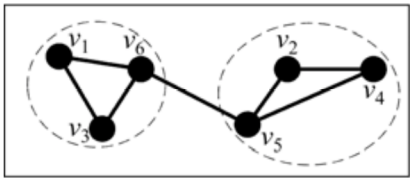
فرض کنید که یک درخت binary و n راسی داریم که در هر راس آن یک عدد از میان 1 تا n نوشته شده است. می‌خواهیم اعداد بر روی این درخت را به گونه‌ای جابه‌جا کنیم که درخت نهایی یک binary search tree باشد(اعداد واقع در زیر درخت سمت راست و چپ هر راس به ترتیب از عدد آن راس بزرگتر و کوچکتر باشند). در هر مرحله می‌توانیم اعداد دو راس مجاور را با یکدیگر جابه‌جا کنیم. برای حل این مساله یک heuristic ارائه دهید و admissible و consistent بودن آن را اثبات کنید.

Genetic

در این قسمت، با استفاده از الگوریتم ژنتیک، به دنبال یافتن یک پاسخ خوب برای یک مسئله هستیم. در این مسئله، هدف ما پیدا کردن گره‌هایی در یک گراف است که درون گروه‌های خود ارتباطات قوی‌تری دارند. به طور دقیق‌تر، ما به دنبال شناسایی گروه‌هایی از این گره‌ها هستیم که بین اعضای گروه، ارتباطات چگال‌تری وجود دارد. توجه داشته باشید که در این مسئله، هر گره فقط به یک گروه تعلق خواهد داشت. به عبارت دیگر امکان اینکه یک گره به بیش از دو گروه تعلق داشته باشد وجود ندارد. همچنین امکان اینکه یک گره به هیچ گروهی تعلق نداشته باشد نیز امکان پذیر نیست.



الف) ژن و کروموزم‌های این مسئله را چگونه تعریف می‌کنید؟
ب) در این مسئله، شکل زیر نشان دهنده یک شبکه شامل ۶ گره و ۲ گروه می‌باشد. این شکل را در مسئله چگونه مشخص می‌کنید؟ (به عبارت دیگر یک encoding از شکل زیر ارائه دهید که بتوانید از آن به عنوان کروموزم در مسئله ژنتیک خود استفاده کنید)



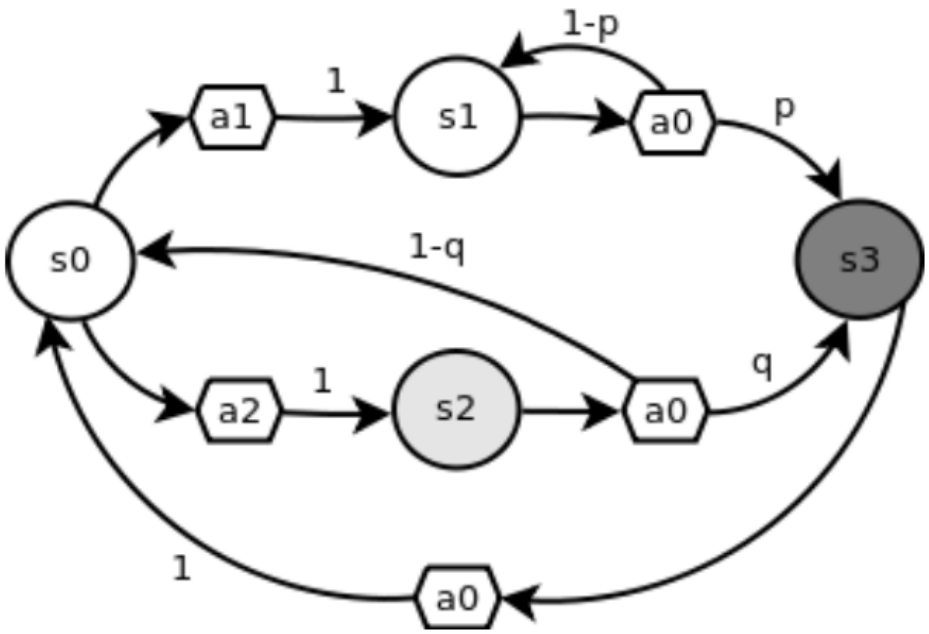
ج) میوتیشن را چگونه تعریف می‌کنید؟ آیا سناریویی وجود دارد که ژن‌های حاصل از میوتیشن نیاز به بازنگری داشته باشند؟ (توجه کنید که میوتیشن باید براساس مدلی باشد که در بخش ب تعریف کرده‌اید)
د) کراس‌اور را چگونه تعریف می‌کنید؟ آیا سناریویی وجود دارد که ژن‌های حاصل از کراس‌اور نیاز به بازنگری داشته باشند؟ (توجه کنید که کراس‌اور باید براساس مدلی باشد که در بخش ب تعریف کرده‌اید)
ه) یک fitness function برای ارزیابی خوب بودن این مسئله ارائه دهید. (امتیازی)

MDP

سوال اول

تصویر زیر بیانگر مسئله MDP افق بینهایت m و با پارامترهای نرخ تخفیف $\gamma \in [0, 1)$ می‌باشد. در این تصویر، state ها با دایره و action ها با شش ضلعی نمایش داده شده اند. عدد نمایش داده شده برروی یال‌های جهت دار، بیانگر احتمال آن انتقال (transition probability) می‌باشد؛ به عنوان مثال $P(s3|s2, a0) = q$. یال‌های نمایش داده نشده، بیانگر احتمال صفر می‌باشند؛ به عنوان مثال $P(s0|s0, a0) = 0$. پاداش در رسیدن به وضعیت s3 برابر با ۱۰ و در رسیدن به وضعیت s2 برابر با ۱ می‌باشد و در غیر این دو وضعیت پاداش صفر است. همچنین $p, q \in [0, 1]$. با توجه به این موارد به سوالات مطرح شده پاسخ بدهید.

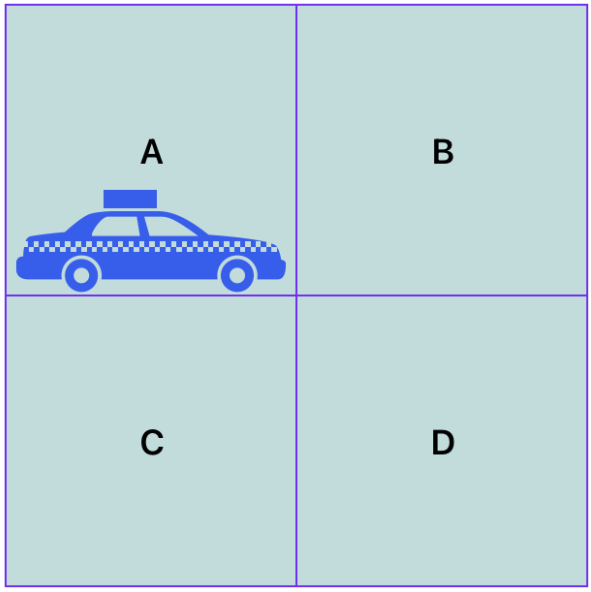
- تمام پالیسی های ممکن برای m را ذکر کنید.
- معادله بیانگر value function بهینه برای هر وضعیت را بنویسید. $(V^*(s_0), V^*(s_1), V^*(s_2), V^*(s_3))$
- آیا هیچ مقداری برای پارامتر p وجود دارد که به ازای تمام $q \in [0, 1]$ و $\gamma \in [0, 1)$، داشته باشیم $\pi^*(s_0) = a2$ ؟ توضیح دهید.
- آیا هیچ مقداری برای پارامتر q وجود دارد که به ازای $p > 0$ و $\gamma \in [0, 1)$، داشته باشیم $\pi^*(s_0) = a1$ ؟ توضیح دهید.
- با استفاده از $\gamma = 0.9$ و $p = q = 0.25$، V^* و π^* را برای تمامی state ها محاسبه کنید. میتوانید از معادلات بخش دوم و یا value iteration استفاده کنید. خطای $\varepsilon = 10^{-3}$ بین V^* و V^t قابل پذیرش است.



سوال دوم

موقعیت زیر را درنظر بگیرید:

- فرض کنید شما یک راننده تاکسی در شهری با چهار مکان A، B، C و D هستید. شما می توانید در هر مکانی مسافران را سوار و پیاده کنید. شما برای هر سفر موفق، بسته به فاصله مبدا و مقصد، مبلغ ثابتی کسب می کنید. شما همچنین برای هر مایلی که رانندگی می کنید، چه با مسافر یا بدون مسافر، هزینه ای متحمل می شوید. شما می توانید انتخاب کنید که در هر مکانی بمانید و منتظر مسافر باشید یا به مکان همسایه رانندگی کنید و در آنجا به دنبال مسافر بگردید. احتمال پیدا کردن مسافر در هر مکان متفاوت است و ممکن است در طول زمان تغییر کند. هدف شما این است که سود کل مورد انتظار خود را در یک روز کاری به حداکثر برسانید.

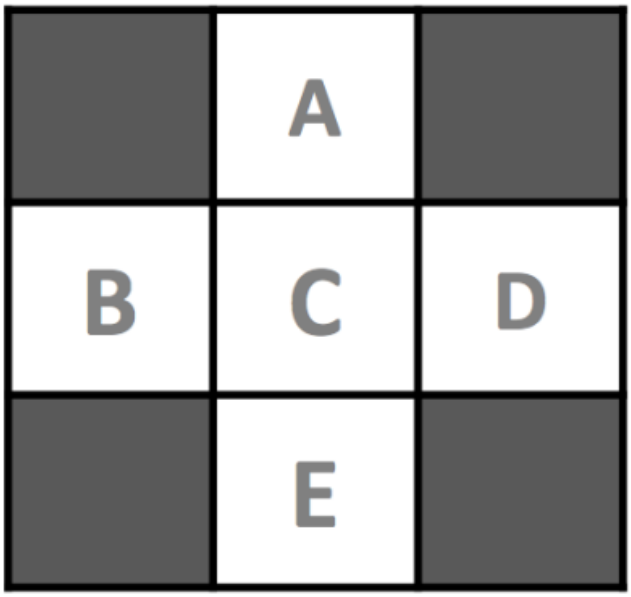


با مشخص کردن اجزای زیر، این مسئله را به عنوان یک MDP مطرح کنید:

1. فضای حالت: حالت های احتمالی که می توانید در آن باشید چیست؟
2. فضای عمل: اقدامات ممکنی که می توانید در هر وضعیت انجام دهید چیست؟
3. مدل انتقال: با توجه به یک عمل، احتمال انتقال از یک حالت به حالت دیگر چقدر است؟
4. تابع پاداش: پاداش (یا هزینه) فوری برای هر جفت حالت-عمل چقدر است؟

RL

فرض کنید یک ربات، در گریدورلد (Gridworld) زیر در حال جمع‌آوری اطلاعات است. او از یک استیت دلخواه شروع می‌کند و با انجام دادن اکشن‌های تصادفی، و با توجه به جایزه‌ای (reward) که به دست می‌آورد، سعی می‌کند به شناخت کافی از این محیط برسد و پالیسی بهینه را به دست بیاورد. هم‌چنین برای انجام این کار از الگوریتم Q-Learning که در درس با آن آشنا شده‌اید استفاده می‌کند.



توجه کنید که در این environment، پنج استیت (A, B, C, D, E) وجود دارد و در هر استیت ایجنت می‌تواند یکی از اکشن‌های West، East، Up و یا South را انجام دهد.

الف)

فرض کنید ایجنت چهار اکشن زیر را به صورت تصادفی در این محیط انجام می‌دهد:

1. از استیت B با انجام اکشن East به استیت C می‌رود، و 2 واحد جایزه می‌گیرد.
2. از استیت C با انجام اکشن South به استیت E می‌رود، و 4 واحد جایزه می‌گیرد.
3. از استیت C با انجام اکشن East به استیت A می‌رود، و 6 واحد جایزه می‌گیرد.
4. از استیت B با انجام اکشن East به استیت C می‌رود، و 2 واحد جایزه می‌گیرد.

بعد از انجام هر مرحله، Q-Table را رسم کنید (در ابتدا تمامی مقادیر جدول برابر صفر هستند) و نحوه‌ی آپدیت شدن جدول را نشان دهید. مقدار learning rate را برابر 0.5 و مقدار discount factor را برابر 1 در نظر بگیرید.

ب)

یک نسخه از الگوریتم Q-Learning با روش Epsilon-greedy را در نظر بگیرید که به جای استفاده از پالیسی استخراج شده از Q-Table فعلی، از یک پالیسی ثابت استفاده می‌کنیم و با احتمال اپسیلون هنوز هم

اکتشاف را انجام می دهیم. اگر این پالیسی ثابت بهینه باشد، عملکرد این الگوریتم چگونه با Q-Learning Epsilon-greedy عادی مقایسه می شود؟